# Birkhoff normalization

Broer, H.; Hoveijn, I.; Lunter, G.; Vegter, G.

*Published in:*
EPRINTS-BOOK-TITLE



*Publication date:*
2003

# 4 Birkhoff normalization

*The Birkhoff normal form procedure is a widely used tool for approximating a Hamiltonian systems by a simpler one. This chapter starts out with an introduction to Hamiltonian mechanics, followed by an explanation of the Birkhoff normal form procedure. Finally we discuss several algorithms for computing the normal form.*

## 4.1 Introduction

Hamiltonian systems with more than 1 degree of freedom are difficult to analyze directly.[1]. The Birkhoff procedure is an iterative procedure for constructing a coordinate transformation that normalizes the system, so that it has extra conserved quantities or *integrals*, and associated *symmetries*. Using these, the system can be reduced to fewer degrees of freedom, see e.g. [AM78, CS85], making the analysis more feasible.

The coordinate transformation resulting from the Birkhoff procedure is a *formal* transformation, which need not converge to an analytic function. The formal series can always be lifted to a $\mathbb{C}^\infty$ transformation, but generically there does not exist a lifting that actually conjugates the original system and the normalized one, e.g. see [BT89]. This ties in with the fact that Hamiltonian systems with at least two degrees of freedom generically are nonintegrable, whereas Birkhoff normalized systems do; see also [Ito89].

Hence, the symmetries and integrals obtained are only *approximate*, up to flat perturbations. For small excitations these perturbations are extremely small, so that integral curves of the system stay close to those of the normalized system for a long time. For two degrees of freedom systems even more is true: KAM tori then prevent the occurrence of Arnol'd diffusion, so that solution curves stay close to the integrable system's tori for all time; see also the introduction to Chap. 2.

Birkhoff's original result [Bir50] dealt with Hamiltonians with a so-called nonresonant quadratic part, close to an elliptic equilibrium. For such a system on, say, a $2n$ dimensional phase space, Birkhoff's method yields $2n$ new coordinates, of which $n$ are integrals of motion, and $n$ are associated *cyclic variables*, each

[1] But see [Sim96].

H. Broer, I. Hoveijn, G. Lunter, and G. Vegter: LNM 1806, pp. 71–84, 2003.

living on a 1-sphere. The system is independent of these $n$ variables, i.e., the system has a $\mathbb{T}^n$ torus symmetry.

Birkhoff's ideas have since been generalized considerably. It is used in the neighborhood of non-elliptic equilibria, around periodic trajectories or invariant tori, and also at resonances. In this work, we are interested in *resonant* Hamiltonians close to *equilibrium*. In this more general situation, Birkhoff's procedure still yields integrals, albeit not as many as before. This is the situation in which we apply the reduction methods of Chaps. 2 and 3; see the respective introductions for details.

This chapter is organized as follows. The first section introduces some aspects of Hamiltonian mechanics, the most important one being the exponential formula for coordinate changes. Then we state and prove Birkhoff's result in the versions we use in Chaps. 2 and 3. The last part is devoted to algorithms for computing the Birkhoff normal form.

## 4.2 Introduction to Hamiltonian mechanics

The aim of this section is to introduce the notions required to state and prove the Birkhoff normal form theorem, and also to introduce the mathematics of Hamiltonian mechanics. For more thorough treatments see [Arn89, MH92, Sja90].

In Hamiltonian mechanics there is an isomorphism[2] between (Hamiltonian) vector fields and the corresponding Hamiltonian function, or *Hamiltonian*, $H$. This isomorphism is given by the *symplectic structure*. Both the the Hamiltonian function and the associated vector field may be transformed by changes of coordinates. The *symplectic* coordinate transformations are, by definition, those that respect the isomorphism. This means that within the class of symplectic transformations, computations on the level of vector fields can also be done on the level of the Hamiltonian functions, which is much easier.

Without much motivation, we now give the basics of Hamiltonian mechanics, and refer to [Arn89, MH92, Sja90] for more details. Let $R = \mathbb{R}^{2n}$ be an even-dimensional space. The coordinates on this space come in pairs $p_1, q_1, \ldots, p_n, q_n$, where the variable $p_i$ is called the *momentum variable* conjugate to the *configuration variable* $q_i$. Heuristically, $q_i$ is a position coordinate, and $p_i$ the associated momentum or 'velocity'. Let $H$ be a function on $R$. Associated to it is a vector field $\boldsymbol{X}_H$ defined by

$$\boldsymbol{X}_H = \sum_i \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i}. \tag{4.1}$$

Here $\frac{\partial}{\partial q_i}$ and $\frac{\partial}{\partial p_i}$ are the constant vector fields in the $q_i$ and $p_i$-directions, respectively. The time-$t$ flow of this vector field is denoted by $\boldsymbol{X}_H^t$. For the derivative of an arbitrary function $F$ in the direction of $\boldsymbol{X}_H$ we write $\boldsymbol{X}_H F$ or $\{H, F\}$:

[2] This holds except for possible global obstructions; on $\mathbb{R}^{2n}$ there are none.

$$\{H, F\} \;:=\; \boldsymbol{X}_H F \;:=\; \frac{\partial}{\partial t} F \circ \boldsymbol{X}_H^t|_{t=0} \;=\; \sum_i \frac{\partial H}{\partial p_i}\frac{\partial F}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial F}{\partial p_i}.$$

The function $\{H, F\}$ is called the *Poisson-bracket* of $H$ and $F$. It is bilinear and antisymmetric, satisfies the Jacobi-identity, and is isomorphic to the ordinary Lie-bracket on vector fields via $\boldsymbol{X}$, that is:

$$\boldsymbol{X}_{\{H,F\}} = [\boldsymbol{X}_H, \boldsymbol{X}_F]. \tag{4.2}$$

Since applying the Poisson-bracket to homogeneous elements yields such elements again, (4.2) induces the structure of a *graded Lie algebra* [Bro79, Bro81] when the base ring and the vector field module are 'chopped up' into homogeneous parts. This structure will be used later.

Definition (4.1) of $\boldsymbol{X}_H$ depends on the coordinates $p_i$ and $q_i$, which is unpleasant when transforming coordinates. For this reason, $\boldsymbol{X}_H$ is usually defined in a coordinate-free manner as the unique vector field satisfying

$$\mathrm{d}H = \omega(\cdot, \boldsymbol{X}_H),$$

where $\omega$ is a nondegenerate differential 2-form, called the *symplectic form*. In the coordinates used above it assumes the form $\omega = \sum_i \mathrm{d}p_i \wedge \mathrm{d}q_i$. Conversely, any nondegenerate closed 2-form $\omega$ can locally be written in this way; this is Darboux' theorem. Coordinate transformations with respect to which $\omega$ is equivariant are called *symplectic*, and also leave the Hamiltonian differential equations (4.1) invariant.

It is easy to see that the flow $\boldsymbol{X}_F^t$ of an arbitrary Hamiltonian $F$ is a coordinate transformation; indeed, $\boldsymbol{X}_F^{-t}$ is its inverse. It turns out that such transformations are symplectic, see e.g. [Arn89, p. 204]. It is therefore natural to try to simplify a Hamiltonian system $H$ by conjugating it with a symplectic transformation generated by another Hamiltonian $F$. The function $H \circ \boldsymbol{X}_F^t$, as a function of $t$, satisfies the differential equation

$$\frac{\partial}{\partial t} H \circ \boldsymbol{X}_F^t = \{F, H \circ \boldsymbol{X}_F^t\}. \tag{4.3}$$

We now define $\mathrm{ad}_F$ to be the operator $\mathrm{ad}_F : H \mapsto \{F, H\}$. This is a linear operator, whence the solution to (4.3) is

$$H \circ \boldsymbol{X}_F^t = \exp(t\,\mathrm{ad}_F)H := H + t\,\mathrm{ad}_F(H) + t^2\frac{1}{2!}\,\mathrm{ad}_F(\mathrm{ad}_F(H)) + \cdots \tag{4.4}$$

(See also [Grö67]). This formula is behind the proof of the Birkhoff normal form theorem.

**Remark 4.1.** *(Formal series)* We denote by $\mathcal{H}_i \subseteq R$ the subspace consisting of homogeneous elements of degree $i$. If $F$ has no linear part, then $\mathrm{ad}_F$ maps $\mathcal{H}_i$ into $\mathcal{H}_i \oplus \mathcal{H}_{i+1} \oplus \cdots$, implying that (4.4) converges to a formal power series. (If $F$ has a linear part, then $\boldsymbol{X}_F^t$ corresponds to a shift of the origin, and conjugation of formal power series with such transformations makes no sense.)

## 4.3 Birkhoff normal form theorem

The idea of Birkhoff's normal form procedure is to conjugate a Hamiltonian system $H$ with the flow of another Hamiltonian $F$, in order to simplify it. It is an iterative procedure, normalizing the system degree by degree. The end result is a Hamiltonian $H'$ in normal form, which has a circle or torus symmetry, and associated conserved quantities.

The Birkhoff procedure results in a formal power series, and any $C^\infty$ transformation with this power series as Taylor series forms a conjugation between $H$ and $H'$, modulo a flat perturbation; see the introduction for more remarks. Although our interest is in practical computations, we below give Birkhoff's result in the formal power series setting, which allows for a clear formulation.

We restrict our attention to the case of normalizing around an equilibrium, which amounts to requiring that the Hamiltonian has vanishing linear part. In that case, the change in a Hamiltonian system $H$, when subjected to a coordinate change generated by $F$ is to first order described by $\operatorname{ad}_{H_2}(F)$. Here $H_2$ is the quadratic part of $H$, and $\operatorname{ad}_{H_2}$ is the associated *adjoint operator*. It is just the derivative of $F$ along the vector field associated to $H_2$. The image of $\operatorname{ad}_{H_2}$ describes, to great extent, the normal form to which $H$ can be transformed. In particular, if $\operatorname{ad}_{H_2}$ is semisimple, then $\ker \operatorname{ad}_{H_2}$ complements $\operatorname{im} \operatorname{ad}_{H_2}$ and, if $F$ is chosen appropriately, the transformed system $H'$ will lie in the kernel of $\operatorname{ad}_{H_2}$. (The equation that $F$ has to satisfy is called the *adjoint equation*). This means that $H'$ is constant under the flow of $H_2$, that is, the flow of $H_2$ is a symmetry of $H'$. Conversely, since $\operatorname{ad}_{H'}(H_2) = -\operatorname{ad}_{H_2}(H') = 0$, we see that $H_2$ is constant under the flow of $H'$; indeed, $H_2$ is the conserved quantity associated to the symmetry. If often happens that more functionally independent quantities are in $\ker \operatorname{ad}_{H_2}$, leading to more symmetries; see Corollary 4.3.

**Theorem 4.2.** *Let $H$ be a Hamiltonian in $\mathbb{R}[[\underline{x}]]$ without linear part, and write it as*

$$H = H_2 + H_3 + \cdots$$

*where $H_i \in \mathcal{H}_i$ is the homogeneous part of degree $i$; see remark 4.1. Let $\mathcal{G}_i \subseteq \mathcal{H}_i$ be linear subspaces such that $\mathcal{G}_i + \operatorname{Im} \operatorname{ad}_{H_2} = \mathcal{H}_i$. Then there exists a formal symplectic power series transformation $\Phi$ such that*

$$H \circ \Phi = H_2 + \widetilde{H}_3 + \widetilde{H}_4 + \cdots$$

*where $\widetilde{H}_i \in \mathcal{G}_i$ $(i = 3, 4, \ldots)$.*

**Proof:** [Tak74c, Arn89, Bro79, Bro81, Mee85, MH92] The proof is by induction. Assume $\phi_i$ normalizes $H$ up to order $i$, so that the partly normalized Hamiltonian is of the form

$$H \circ \phi_i = H_2 + \widetilde{H}_3 + \cdots + \widetilde{H}_i + H'_{i+1} + \cdots$$

where $\widetilde{H}_j \in \mathcal{G}_j$, $j = 3, \ldots, i$. Choose $F_{i+1} \in \mathcal{H}_{i+1}$ such that $\widetilde{H}_{i+1} := H'_{i+1} - \operatorname{ad}_{H_2}(F_{i+1}) \in \mathcal{G}_{i+1}$, and define $\phi_{i+1} := \phi_i \circ \boldsymbol{X}^1_{F_{i+1}}$. By (4.4) we get

$$\begin{aligned} H \circ \phi_{i+1} &= H_2 + \widetilde{H}_3 + \cdots + \widetilde{H}_i + H'_{i+1} + \cdots + \mathrm{ad}_{F_{i+1}}(H_2) + \mathrm{ad}_{F_{i+1}}(\widetilde{H}_3) + \cdots \\ &= H_2 + \widetilde{H}_3 + \cdots + \widetilde{H}_i + (H'_{i+1} - \mathrm{ad}_{H_2}(F_{i+1})) + \cdots \\ &= H_2 + \widetilde{H}_3 + \cdots + \widetilde{H}_i + \widetilde{H}_{i+1} + H''_{i+2} + \cdots . \end{aligned}$$

(Here we used that $\mathrm{ad}_{H_2}$ maps $\mathcal{H}_i$ to itself.) So $\phi_{i+1}$ normalized $H$ up to order $i+1$. Since $\phi_i$ and $\phi_{i+1}$ coincide up to terms of order $i$, the sequence $(\phi_i)_{i=1}^{\infty}$ converges as a formal power series. ∎

To see how theorem 4.2 implies the extra symmetries, we also formulate Birkhoff's original result:

**Corollary 4.3.** *Let $H \in \mathbb{R}[[\underline{x}]]$ be a Hamiltonian without linear part, and with quadratic part*

$$H_2 = \omega_1(q_1^2 + p_1^2) + \cdots + \omega_n(q_n^2 + p_n^2).$$

*Assume that the non-resonance condition holds: For all nonzero vectors $(k_1, \ldots, k_n)$ in $\mathbb{Z}^n$ we have $k_1\omega_1 + \cdots + k_n\omega_n \neq 0$. Then there exists a formal symplectic power series transformation $\Phi$ such that*

$$H \circ \Phi = H'(L_1, \ldots, L_n),$$

*where $L_i = q_i^2 + p_i^2$. The system $H'$ has a $\mathbb{T}^n$ torus symmetry.*

The $L_i$ are the 'radial' part of so-called symplectic Hamiltonian polar coordinates $(L_i, \phi_i)$. The system $H'$ is independent of the $\phi_i$, so that the $\mathbb{T}^n$ torus action

$$\zeta : \quad \mathbb{T}^n \times R \to R : \quad (\rho_i, L_i, \phi_i) \mapsto (L_i, \phi_i + \rho_i)$$

is a $\mathbb{T}^n$-symmetry of the system. The $L_i$ are conserved by the flow, since by Hamilton's equations $\frac{\partial}{\partial t} L_i = \frac{\partial H}{\partial \phi_i} = 0$.

**Proof** of the corollary: It is convenient to use complex coordinates $z_i = q_i + \mathrm{i}p_i$, then $H_2 = \omega_1 z_1 \bar{z}_1 + \cdots$, and it is easily checked that monomials in these variables are eigenfunctions of the adjoint operator:

$$\mathrm{ad}_{H_2}(z_1^{\alpha_1}\bar{z}_1^{\beta_1}\cdots z_n^{\alpha_n}\bar{z}_n^{\beta_n}) = (\omega_1(\alpha_1 - \beta_1) + \cdots + \omega_n(\alpha_n - \beta_n))\,(z_1^{\alpha_1}\bar{z}_1^{\beta_1}\cdots z_n^{\alpha_n}\bar{z}_n^{\beta_n}).$$

Using the non-resonance condition, it is seen that the eigenvalues are zero exactly for elements of the algebra generated by the functions $L_1 = z_1\bar{z}_1, \ldots, L_n = z_n\bar{z}_n$. Taking $\mathcal{G}_i$ to be the subspace of degree-$i$ elements of this algebra, the result follows directly from the conclusion of theorem 4.2. ∎

**Remark 4.4.** *(Symmetric normalizing)* Often the Hamiltonian $H$ is invariant with respect to some symmetry group $\Gamma$. We assume that $\Gamma$ is compact and acts linearly on the phase space. We also assume that $\Gamma$ respects the symplectic structure, i.e., $\{F, G\} \circ \gamma = \{F \circ \gamma, G \circ \gamma\}$ for all $\gamma \in \Gamma$. Then, in theorem 4.2 we may restrict to the space of $\Gamma$-invariant Hamiltonians, and $\phi$ can be chosen to

be $\Gamma$-equivariant, as the image of $\mathrm{ad}_{H_2}$ restricted to $\Gamma$-invariant Hamiltonians is element-wise $\Gamma$-invariant. If $\Gamma$ *anti*-commutes with the Poisson-bracket, as happens if $\Gamma$ is the time-reversal symmetry, then above remarks still hold true but the argument involves a number of minus-signs.

**Remark 4.5.** *(Normalizing a family)* In our application, the Hamiltonian $H$ depends on small *parameters* $u_i$. In particular, the quadratic part depends on the $u_i$, so that $\ker \mathrm{ad}_{H_2}$ is also dependent on the parameters. However, it follows from the proof that one can find $F_i$, with rational functions of the $u_i$ as coefficients, such that $\widetilde{H}_i \in \ker \mathrm{ad}_{H_2^0}$, where $H_2^0$ is the quadratic part of $H$ for $u_1 = u_2 = \cdots = 0$. Moreover, the coefficients of the $\widetilde{H}_i$ are also rational functions in the $u_i$. It can be shown that the coefficient-values for which a term $F_i$ or $\widetilde{H}_i$ is singular (i.e., where its coefficients have a pole) are generally dense in parameter-space, and correspond to high-order resonances in $H_2$. If one normalizes only to finite order, a full neighborhood of the origin is free from such singularities.

**Remark 4.6.** *(Generating function)* The normalizing transformation $\phi$ is constructed as a composition $\boldsymbol{X}^1_{F_3} \circ \boldsymbol{X}^1_{F_4} \circ \cdots$, with infinitely many components $\boldsymbol{X}^1_{F_i}$ and $F_i \in \mathcal{H}_i$. The resulting transformation may be written in the form $\boldsymbol{X}^1_{F'}$, where $F' = F'_3 + F'_4 + \cdots$, and again $F'_i \in \mathcal{H}_i$. The relation between the $F_i$ and the $F'_i$ is given by the Campbell-Baker-Hausdorff formula; see [DF76].

### 4.3.1 Semisimple quadratic part, and resonance

In this section, Theorem 4.2 is applied to the case where $H$ has a semisimple quadratic part without or with just one resonance.

Instead of coordinates $p_i, q_i$ we use complex $\mathbb{R}$-linearly independent coordinates $z_i, \bar{z}_i$ defined by $z_i := q_i + \mathrm{i}p_i$, $\bar{z}_i := q_i - \mathrm{i}p_i$. By assumption $H_2$ is semisimple, and in these coordinates this means that the matrix $X_{H_2}$ is diagonalizable. Transforming coordinates (symplectically), we may suppose that $H_2$ is of the form[3]

$$H_2 = \mathrm{i}\omega_1 z_1 \bar{z}_1 + \cdots + \mathrm{i}\omega_n z_n \bar{z}_n. \tag{4.5}$$

In complex coordinates, $\omega = \mathrm{d}z \wedge \mathrm{d}\bar{z}$. It follows that since $X_{H_2}$ is semisimple, so is the operator $\mathrm{ad}_{H_2}$. In fact its eigenfunctions are the monomials, in complex coordinates:

$$\begin{aligned}\mathrm{ad}_{H_2}(z_1^{\alpha_1}\bar{z}_1^{\beta_1}\cdots z_n^{\alpha_n}\bar{z}_n^{\beta_n}) &= \left(\sum_{i=1}^n \omega_i\beta_i - \omega_i\alpha_i\right) z_1^{\alpha_1}\bar{z}_1^{\beta_1}\cdots z_n^{\alpha_n}\bar{z}_n^{\beta_n} \\ &= (\omega, \beta - \alpha) z_1^{\alpha_1}\bar{z}_1^{\beta_1}\cdots z_n^{\alpha_n}\bar{z}_n^{\beta_n}.\end{aligned}$$

[3] The transformation from coordinates $\{p_1, p_2, \ldots, q_1, q_2, \ldots\}$ to $\{z_1, z_2, \ldots, \bar{z}_1, \bar{z}_2, \ldots\}$ is a symplectic transformation with *multiplier* 2i. This accounts for the factor i in (4.5). See also Chap. 2, remark 2.3.

The kernel of $\mathrm{ad}_{H_2}$ is spanned by the monomials with exponents $(\alpha, \beta) \in \mathbb{N}^{2n}$ satisfying $(\beta - \alpha, \omega) = 0$. This set of exponents forms a semi-lattice (a set that is closed under addition, not under subtraction), therefore $\ker \mathrm{ad}_{H_2}$ is an algebra. We consider two cases, namely that the components of $\omega$ are independent over $\mathbb{Q}$ (the non-resonant case), and that there exists essentially one relation over $\mathbb{Q}$.

**Lemma 4.7.** *Let $H_2$ be of the form (4.5). Define $A := \{(\alpha, \beta) \in \mathbb{N}^{2n} : (\beta - \alpha, \omega) = 0\}$. Let $a_i$ be the exponent vector associated to the monomial $z_i \bar{z}_i$.*

*a) If the $\omega_i$ are independent over $\mathbb{Q}$, then $A = \mathbb{N}\{a_1, \ldots, a_n\}$ and $\ker \mathrm{ad}_{H_2}$ is generated, as an algebra, by*

$$\{z_1 \bar{z}_1, \ldots, z_n \bar{z}_n\}.$$

*b) If the $\omega_i$ obey one (up to scalar multiplication) non-trivial relation $\sum_i (\nu_i - \mu_i)\omega_i = 0$ with $\nu_i, \mu_i \in \mathbb{N}$ and $\nu_i \mu_i = 0$, then $A = \mathbb{N}\{a_1, \ldots, a_n, (\mu, \nu), (\nu, \mu)\}$ and $\ker \mathrm{ad}_{H_2}$ is generated, as an algebra, by*

$$\{z_1 \bar{z}_1, \ldots, z_n \bar{z}_n, z_1^{\nu_1} \bar{z}_1^{\mu_1} \cdots z_n^{\nu_n} \bar{z}_n^{\mu_n}, z_1^{\mu_1} \bar{z}_1^{\nu_1} \cdots z_n^{\mu_n} \bar{z}_n^{\nu_n}\}.$$

The use of this lemma lies in the fact that, for semisimple operators like $\mathrm{ad}_{H_2}$ we have

$$\ker \mathrm{ad}_{H_2} \oplus \mathrm{Im}\, \mathrm{ad}_{H_2} = \bigoplus_{i=3}^{\infty} \mathcal{H}_i.$$

In other words, for the $\mathcal{G}_i$ in theorem 4.2 we can take $\ker \mathrm{ad}_{H_2} \cap \mathcal{H}_i$.

**Remark 4.8.** *(Unique normal form)* If $H_2$ is nonresonant, this choice of $\mathcal{G}_i$ is best possible and the normal form is unique. In case of resonance, a normal form with $\mathcal{G}_i = \ker \mathrm{ad}_{H_2}$ is not unique; see [CS85, SvdM92, Mee85]. See also Sect. 4.3.2, where this non-uniqueness is exploited in the case of the $1:2$-resonance.

More specifically, for a singly-resonant system the statement is like this:

**Proposition 4.9.** *Let $H$ be a Hamiltonian with $n$ degrees of freedom. Suppose that the quadratic part $H_2$ is of the form*

$$H_2 = \mathrm{i}\omega_1 z_1 \bar{z}_1 + \mathrm{i}\omega_2 z_2 \bar{z}_2 + \cdots + \mathrm{i}\omega_n z_n \bar{z}_n$$

*with $\frac{\omega_2}{\omega_1} = \frac{p}{q}$, $\gcd(p, q) = 1$, $q > 0$, and suppose there are no further relations over $\mathbb{Q}$ between the $\omega_i$. Then there exists a formal symplectic coordinate transformation $\phi$ such that*

$$H \circ \phi = \begin{cases} H_2 + f(z_1 \bar{z}_1, \cdots, z_n \bar{z}_n, z_1^p \bar{z}_2^q, \bar{z}_1^p z_2^q) & \text{if } p > 0, \\ H_2 + f(z_1 \bar{z}_1, \cdots, z_n \bar{z}_n, z_1^{-p} z_2^q, \bar{z}_1^{-p} \bar{z}_2^q) & \text{if } p < 0. \end{cases}$$

**Remark 4.10.** The normalized singly-resonant system has an $n-1$ torus-symmetry $\mathbb{T}^{n-1} \ni (\phi_1, \phi_3, \ldots, \phi_n)$, with action

$$\zeta : (\phi, z) \mapsto (\mathrm{e}^{2\pi \mathrm{i} q \phi_1} z_1, \mathrm{e}^{2\pi \mathrm{i} p \phi_1} z_2, \mathrm{e}^{2\pi \mathrm{i} \phi_3} z_3, \ldots, \mathrm{e}^{2\pi \mathrm{i} \phi_n} z_n).$$

### 4.3.2 Second normalization

In the proof of Theorem 4.2, the functions $F_m$ were unique up to terms in $\ker \mathrm{ad}_{H_2}$. Part of this non-uniqueness reflects the fact that Hamiltonians have infinitely many invariants under symplectic transformations.[4] However, different choices of $F_m$ *do* change the final normal form, since a term in $\ker \mathrm{ad}_{H_2}$ may still give a contribution to the sum (4.4) via terms of higher order than $H_2$ in $H$. Pursuing this idea leads to *unique normal forms*. The result for systems in $1:2$ resonance can be found in [SvdM92]. In Chap. 3 the following result, which is a consequence of the unique $1:2$-resonant normal form, is used. We use the following abbreviations for the fundamental invariants:

$$\rho_1 = z_1\bar{z}_1, \quad \rho_2 = z_2\bar{z}_2, \quad \psi = \frac{1}{2}(z_1^p\bar{z}_2^{|q|} + \bar{z}_1 z_2^{|q|}), \quad \chi = \frac{1}{2\mathrm{i}}(z_1^p\bar{z}_2^{|q|} - \bar{z}_1^p z_2^{|q|}).$$

**Proposition 4.11.** *Suppose $\omega = p:q = 1:2$ or $1:-2$, and suppose $H$ is a Hamiltonian with 2 degrees of freedom with quadratic part*

$$H_2 = i\rho_1 + i\omega\rho_2.$$

*Let $\phi_1$ be the normalizing transformation of Proposition 4.9, and suppose that the coefficients of $\psi$ and $\chi$ in $H \circ \phi_1$ do not both vanish. Then there exists a symplectic transformation $\phi_2$ such that*

$$H \circ \phi_2 = H_2 + f(\rho_1, \rho_2, \psi).$$

**Proof:** (Sketch) By Proposition 4.9 there exists a $\phi_1$ that brings $H$ in the form

$$H_2 + f_0(\rho_1, \rho_2, \psi, \chi). \tag{4.6}$$

The second normalization is done by applying successive coordinate transformations $X^1_{F_s}$ for increasing $s$, similar to the ordinary Birkhoff procedure. The homogeneous degree-$s$ generator $F_s$ are now required to lie in $\ker \mathrm{ad}_{H_2}$. Jacobi's identity says

$$\{H_2, \{F, G\}\} + \{F, \{G, H_2\}\} + \{G, \{H_2, F\}\} = 0,$$

and it follows that $\ker \mathrm{ad}_{H_2}$ is closed under the operation of taking Poisson-brackets. Hence, by (4.4), transformations generated by the $F_s$ do not disturb the general shape (4.6) of the normal form. To see how these transformations act on $H \circ \phi_1$, we look at the first nonzero homogeneous part with degree higher than $H_2$, which is $a\psi + b\chi$. First apply a linear symplectic transformation $z_1 \mapsto \zeta z_1, z_2 \mapsto z_2$ with $|\zeta| = 1$ to reduce to the case $b = 0$.

[4] For example, let $H(p, q)$ be a 1 d.o.f. system with elliptic equilibrium at the origin such that $H(0, 0) = 0$, and let $A(h)$ be the surface area bounded by the level curve $H = h$, then the Taylor coefficients of $A(h)$ around some $h > 0$ are (*symplectic*, not *dynamic*) invariants of $H$.

Since $\psi^2+\chi^2-\rho_1\rho_2^2=0$, it is possible to write $f_0(\rho_1,\rho_2,\psi,\chi)=f_0'(\rho_1,\rho_2,\psi)+\chi f_0''(\rho_1,\rho_2,\psi)$. It therefore suffices to remove all terms of the form $\rho_1^t\rho_2^u\psi^v\chi$. Taking the Poisson-bracket of the monomials $\rho_1^n\rho_2^m\psi^k$ and $2i\psi$ yields

$$\begin{aligned}\{2\mathrm{i}\psi,\rho_1^n\rho_2^m\psi^k\}&=2\rho_1^{n-1}\rho_2^{m-1}(2m\rho_1-n\rho_2)\chi\psi^k,\\ \{2\mathrm{i}\psi,\rho_1^n\psi^k\}&=-2n\rho_1^{n-1}\chi\psi^k,\\ \{2\mathrm{i}\psi,\rho_2^m\psi^k\}&=4m\rho_2^{m-1}\chi\psi^k.\end{aligned}$$

This shows that indeed all relevant terms can be removed, provided that the coefficients of $\psi$ and $\chi$ in $H\circ\phi_1$ do not vanish simultaneously. ∎

## 4.4 Algorithms for the Birkhoff normal form

An important part of the Birkhoff normal form computation, is the transformation of the Hamiltonian by a symplectic coordinate change. Related methods for vector fields, on which the Hamiltonian versions are based, can be found in e.g. [Tak74c, Bro81]. The oldest methods for transforming coordinates of Hamiltonian systems is the mixed-variable generating function method, also called the Hamilton-Jacobi method or Poincaré-von Zeipel method; see [Arn89, Car81, LL92, How77, RA87, SV85]. This method seems however less suited to computer implementation (see [LL92]).

Today, the most widely-used method for computing Birkhoff normal forms is an algorithm by Deprit [Dep69, DHPR69], known also as the *Lie-triangle algorithm*. Recent expositions are [MH92, LL92, GG78]. Several implementations are available, e.g. see [ASJ93] or [RA87]. Various variations of Deprit's algorithm are known in literature [CR89, Hen70], one of which is particularly fast on Hamiltonians of special form [Hen73].

A third class of methods is based on the exponential formula (4.4), and are (distantly) related to Kolmogorov's quadratically convergent method. These methods have a lower time-complexity than the Lie-triangle algorithms. In [How77, HR84, DF76, Car81], essentially this method is presented, though not in ways that suggest efficient implementations on a computer. In this section we give Deprit's algorithms, and an algorithm based on (4.4), and discuss their computational complexity.

Analyzing the complexity of the algorithms is not straightforward. The algorithms treated below use two main types of algebraic manipulations: calculation of Poisson brackets, and solving the adjoint 'equation' $\mathrm{ad}_{H_2}F\in\mathcal{G}$. The latter is the same for all algorithms, so the number of Poisson-bracket calculations is a better indicator of the algorithm's efficiency. However, the time required for doing one such calculation depends on the number of terms in the factors: The number of terms in the bracket equals, roughly, the product of those in the factors. As a compromise, we chose to count the number of Poisson-bracket calculations performed on *homogeneous* factors. In fact this is measure of complexity is commonly used; see e.g. [Car81, BP97, Hen70].

### 4.4.1 Another formulation of Birkhoff's result

Usually Birkhoff's normal form theorem is formulated as a *perturbation* result, where the Hamiltonian to be normalized depends on a small parameter $\epsilon$, and for $\epsilon = 0$ the Hamiltonian is integrable. The following result is similar to Theorem 4.2. The 'interface' is provided by setting $\mathcal{G} = \mathcal{G}_3 \oplus \mathcal{G}_4 \oplus \cdots$ and having $H_i$ be homogeneous of degree $i+2$, then also $\widetilde{H}_i$ is homogeneous of degree $i+2$, and putting $\epsilon = 1$ in the end is allowed. This procedure is known as 'scaling', and since we use formal power series convergence in $\epsilon$ is not an issue.

**Theorem 4.12.** *Let $H = H_0 + \epsilon H_1 + \epsilon^2 H_2 + \cdots$ be a Hamiltonian, with $H_i \in \mathcal{H}$, and such that $H_0$ has no linear part. Let $\mathcal{G} \subseteq \mathcal{H}$ be such that $\operatorname{Im} \operatorname{ad}_{H_0} \oplus \mathcal{G} = \mathcal{H}$. Then there exists a formal (in $\epsilon$) symplectic coordinate transformation $\phi$ such that*

$$H \circ \phi = H_0 + \epsilon \widetilde{H}_1 + \epsilon^2 \widetilde{H}_2 + \cdots$$

*with $\widetilde{H}_i \in \mathcal{G}$ for $i = 1, 2, \ldots$.*

**Proof:** The proof is analogous to that of Theorem 4.2. ∎

### 4.4.2 Deprit's algorithm

Two versions of Deprit's algorithm are around. Both have as output a generating function $W(\epsilon) = W_0 + \epsilon W_1 + \epsilon^2 W_2 + \cdots$. Here $\epsilon$ is interpreted as a time variable, and the $W_i$ are chosen such that $H \circ X^{\epsilon}_{W(\epsilon)} = \widetilde{H}$, where $X^{\epsilon}_{W(\epsilon)}$ is the time-$\epsilon$ flow of the 'time'-dependent phase flow of the Hamiltonian $W$. The central part of both algorithms is a recipe for computing the composition $H \circ X^{\epsilon}_{W(\epsilon)}$.

**Slow version** The first version of Deprit's algorithm can be summarized as follows. For the non-autonomous version see e.g. [RA87, LL92]; here we stick to the autonomous case.

$$W = \sum_{i=1}^{\infty} \epsilon^{i-1} W_i, \qquad \widetilde{H} = \sum_{i=0}^{\infty} \epsilon^i K_i,$$

$$S_0 = \text{identity operator}, \qquad S_n = \frac{1}{n} \sum_{i=0}^{n-1} \operatorname{ad}_{W_{n-i}} S_i \qquad (n \geq 1),$$

$$K_n = H_n + \frac{1}{n} \sum_{i=0}^{n-1} \left( \operatorname{ad}_{W_{n-i}} K_i + i S_{n-i} H_i \right). \tag{4.7}$$

This algorithm is derived by formally solving the differential equation (in $\epsilon$) for the operator 'composition with $X^{\epsilon}_{W(\epsilon)}$', resulting in the expressions for $S_n$, and calculating what new Hamiltonian corresponds to $H$ in the new variables. See [Car81] for a derivation.

The $S_n$ are operators. On a computer these are implemented as recursive subroutines. Computing $S_n$ requires $2^n - 1$ Poisson-brackets, computing $K_n$ involves $n + \sum_{i=1}^{n-1}(2^{n-i} - 1) = 2^n - 1$ brackets, and computing $\widetilde{H}$ up to $O(\epsilon^{n+1})$ terms takes $2^{n+1} - (n+2)$ brackets. For large $n$ then, this algorithm is slow.

**Fast version** By merging the differential equation for the $S_n$ with the formulas for the Hamiltonian in the new coordinates, Deprit arrives at a faster algorithm with the same functionality. (See [Dep69], or [MH92] for a clear derivation.) It uses a triangular array of intermediate Hamiltonians $H_j^i$, and is summarized as follows:

$$W = \sum_{i=0}^{\infty} \epsilon^i W_i, \qquad H = \sum_{i=0}^{\infty} \epsilon^i H_i^0, \qquad \widetilde{H} = \sum_{i=0}^{\infty} \epsilon^i H_0^i,$$

$$H_j^i = \frac{1}{i}\left((j+1)H_{j+1}^{i-1} + \sum_{k=0}^{j} \mathrm{ad}_{W_k} H_{j-k}^{i-1}\right). \tag{4.8}$$

**Proposition 4.13.** *Algorithm (4.8) performs* $\frac{1}{6}n(n+1)(n+2)$ *Poisson-bracket computations to compute* $\widetilde{H}$ *up to* $O(\epsilon^{n+1})$ *terms.*

**Proof:** Computing $H_j^i$ costs $j+1$ Poisson-brackets, and requires $H_k^{i-1}$ to be known for $k = 0, \ldots, j+1$. To compute $H_0^i$ for $i = 0, \ldots, n$ we therefore need the upper-triangular part of the matrix $H_j^i$ with $i + j \leq n$, computed at a cost of $\sum_{i=1}^{n} \sum_{j=0}^{n-i}(j+1) = \frac{1}{6}n(n+1)(n+2)$ Poisson-brackets. ∎

**Remark 4.14.** *(Efficiency)* It is not necessary to store the entire array $H_j^i$ in memory. By traversing it suitably, it suffices to store two rows.

**Remark 4.15.** *(Homogeneous perturbations)* In [Hen73], Henrard develops a modification of algorithm (4.8). This algorithm has the same asymptotic order as (4.8), however the actual number of Poisson brackets is often lower in practice, and if the perturbation is *homogeneous*, only $n^2$ Poisson bracket computations are necessary. For such Hamiltonians, Henrard's algorithm also outperforms the algorithm of Sect. 4.4.3.

**Remark 4.16.** *(Solving the adjoint equation)* Above algorithms do not constitute methods for computing the normal form yet. What fails is a method to construct $W$. In practice this is an additional calculation that can be done while computing the composition, see e.g. [Lun94]. If diagonalizing complex coordinates are used, this computation is straightforward.

### 4.4.3 The exponential-map algorithm

The idea of this algorithm is to write the normalizing transformation as a composition of transformations, each normalizing the Hamiltonian at a specific order of $\epsilon$. These transformations are the flows of generating Hamiltonians $\epsilon^k W_k$, as in Deprit's case. The final transformation therefore is of the form

$$\phi = \epsilon X^1_{W_1} \circ \epsilon^2 X^1_{W_2} \circ \cdots$$

This should be contrasted to the form of the transformation in Deprit's case,

$$\phi = X^\epsilon_{W_1 + \epsilon W_2 + \epsilon^2 W_3 + \cdots}$$

It turns out that the former coordinate transformation (composed with a Hamiltonian function $H$) can be computed more efficiently than the latter. Our algorithm is based on the proof of the normal form theorem for vector fields in [Tak74c]. It resembles Dragt and Finn's algorithm [DF76], with the difference that we evaluate all operators directly instead of postponing evaluation to the end. This results in a dramatic increase of efficiency. In fact Dragt and Finn's algorithm is not very efficient, asymptotically: in [Car81] the estimate $\exp(\pi\sqrt{2n/3})/(4n\sqrt{3})$ is given for the number of Poisson brackets to compute to order $n$, comparing favorably with (4.7) but not with (4.8).

Howland [How77, HR84] calculates normal forms using a method related to Kolmogorov's quadratically convergent procedure. Quadratic convergence translates to $O(k^2 \log k)$ Poisson-bracket computations, which makes it of the same order as the algorithm proposed in this section. Howland does not show how his method could be implemented, however.

The exponential-map algorithm is based on the proof of Theorem 4.2, and formula (4.4); see algorithm 4.17. Note that $K_0 = H_0$ throughout the algorithm, and is referred to only in the first line of the inner for-loop when $j = 0$: $-\operatorname{ad}_{K_j}(F_i) = -\operatorname{ad}_{H_0}(F_i) = \operatorname{ad}_{F_i}(H_0)$. This value can be computed easily while computing $F_i$. The algorithm can therefore be slightly rewritten so that this ad-computation is not needed, which also removes the need of initializing $K_0$, but for clarity this has not been done. Note also that the $K_i$ is the only storage needed, and memory requirements are less than for Deprit's algorithm by a factor 2.

**Algorithm 4.17.** *(Birkhoff normal form)*
Input: $H_i \in \mathcal{H}$ *for* $i = 0, \ldots, k$, *subspace* $\mathcal{G} \subset \mathcal{H}$ *s.t.* $\operatorname{Im}\operatorname{ad}_{H_0} \oplus \mathcal{G} = \mathcal{H}$
Output: $F_i, \widetilde{H}_i \in \mathcal{H}$ *for* $i = 1, \ldots, k$, *such that*

$$\begin{aligned} &\widetilde{H}_i \in \mathcal{G} \qquad (i = 1, \ldots, k) \\ &(H_0 + \epsilon H_1 + \cdots + \epsilon^k H_k) \circ X^1_{\epsilon F_1} \circ \cdots \circ X^1_{\epsilon^k F_k} = \\ &\qquad = H_0 + \epsilon \widetilde{H}_1 + \cdots + \epsilon^k \widetilde{H}_k + O(e^{k+1}) \end{aligned}$$

Complexity: $O(k^2 \log k)$ *Poisson-brackets are computed.*

Algorithm:

$K_i \leftarrow H_i \qquad (i = 0, \dots, k)$
*For* $i$ *from* 1 *to* $k$, *do the following:*
  *Find* $F_i$ *such that* $\mathrm{ad}_{H_0}(F_i) - K_i \in \mathcal{G}$
  *For* $j$ *from* $k - i$ *down to* 0, *do the following:*
    $F \leftarrow -\mathrm{ad}_{K_j}(F_i)$
    $K_{i+j} \leftarrow K_{i+j} + F$
    $q \leftarrow 2$
    *While* $qi + j \le k$ *do the following:*
      $F \leftarrow \mathrm{ad}_{F_i}(F)/q$
      $K_{qi+j} \leftarrow K_{qi+j} + F$
      $q \leftarrow q + 1$
    *EndWhile*
  *EndFor*
*EndFor*
*Output* $F_i$, *and* $\widetilde{H}_i = K_i$.

**Proof** of the algorithm: We prove correctness by proving invariance, over the outer for-loop, of the following assertion:

(4.9)
$$\begin{cases} (H_0 + \epsilon H_1 + \cdots + \epsilon^k H_k) \circ X^1_{\epsilon F_1} \circ \cdots \circ X^1_{\epsilon^i F_i} = \\ \qquad\qquad K_0 + \epsilon K_1 + \cdots + \epsilon^k K_k + O(\epsilon^{k+1}), \\ K_j \in \mathcal{G}. \qquad (j = 1, \dots, i) \end{cases}$$

After initialization and for $i = 0$, (4.9) is trivially true. For $i = k$ it implies the output condition on $F_i$ and $\widetilde{H}_i$. So we are done if we prove invariance of (4.9).

The inner for-loop computes $(K_0 + \epsilon K_1 + \cdots + \epsilon^k K_k) \circ X^1_{\epsilon^i F_i}$. By (4.4), and using linearity of the ad-operator, the order-$\epsilon^i$ term in this expression is $K_i + \mathrm{ad}_{F_i}(H_0) = K_i - \mathrm{ad}_{H_0}(F_i)$. By the choice of $F_i$, indeed $K_i \in \mathcal{G}$ after this pass through the outer for-loop, and $K_i$ does not change in subsequent passes.

Each pass through the inner for-loop computes $\epsilon^j K_j \circ X^1_{\epsilon^i F_i}$ for some $j$, up to $O(\epsilon^{k+1})$ terms. By (4.4), this is a sum of terms of the form $\epsilon^{qi+j} \mathrm{ad}^q_{F_i}(K_j)/q!$ for $q = 0, \dots, \infty$, and this is just what is computed recursively in the inner While-loop, for $q$-values that contribute to terms of order $k$ or less. By having the counter $j$ counting downwards, the high order $K_j$'s are computed first, eliminating the need of storing the 'new' $K_j$'s in a separate array to prevent overwriting the old ones. This proves invariance of (4.9), and correctness of the algorithm.

The number of Poisson-brackets computed by the algorithm can be estimated as follows. Through one pass of the inner for-loop, at most $(k-j)/i$ brackets are computed. Summing this over $j = 0, \dots, k-i$ this becomes $k(k+1)/(2i) - (i-1)/2$ for the number of brackets computed in the inner for-loop. The total number of brackets is then estimated by

$$\sum_{i=1}^{k} \frac{k(k+1)}{2i} - \frac{i-1}{2} \leq \frac{k(k+1)}{2}\mathrm{Harm}_k - \frac{k(k-1)}{4} = O(k^2 \log k),$$

where $\mathrm{Harm}_k$ is the $k$th harmonic number, proving the asymptotic order of the algorithm. ∎

**Remark 4.18.** *(Normalizing transformation)* As it stands, the algorithm only computes the normal form, not the normalizing coordinate transformation. This transformation is obtained by evaluating $I \circ X^1_{\epsilon^3 F_3} \circ \cdots \circ X^1_{\epsilon^n F_n}$, where $I$ is the identity transformation, which is very similar to the computation of $H \circ X^1_{\epsilon^3 F_3} \circ \cdots$ done by the algorithm. Slight modifications are necessary to reflect the fact that $I$ has linear terms, whereas $H$ is quadratic to leading order. The resulting algorithm takes approximately twice as much time as algorithm 4.17, hence is also $O(n^2 \log n)$.